



ELSEVIER

Contents lists available at ScienceDirect

Injury

journal homepage: www.elsevier.com/locate/injury

Patient-reported outcomes: Which ones are most relevant?

Esther M.M. Van Lieshout*, Mathieu M.E. Wijffels

Trauma Research Unit Department of Surgery, Erasmus MC, University Medical Center Rotterdam, 3000 CA Rotterdam, P.O. Box 2040, The Netherlands

ARTICLE INFO

Article history:

Accepted 22 October 2019

Available online xxx

Keywords:

PRO

Patient-reported outcome

Outcome

Validity

Responsiveness

Reliability

ABSTRACT

The use of patient-reported outcome measures (PROs) in research and clinical use in orthopedic trauma surgery has increased dramatically over the past decades, and this trend will continue in the future. This review aimed to provide insight into the way PROs can best be identified, reviewed, selected, and used in orthopedic trauma studies and patient care. Selection of instruments for a specific (research) question starts with a systematic literature review for identifying the mostly used instruments. From this list, the instruments that are most suitable for the intended use are selected. Readability, usability, and costs should be considered at this stage. For the selected instruments, the next step is to search the literature systematically in order to determine if the measurement properties of these instruments meet the quality standard as outlined by the COSMIN group. Instruments need to be valid for the targeted condition, intervention as well as population. If not available in the primary language of the targeted population, the instrument requires translation according to a state of the art procedure and validation in the new language. Reporting guidelines like the SPIRIT-PRO and CONSORT-PRO are available in order to guide the reporting of PROs in study protocol and outcome reports, respectively. Using Core Outcome Sets improves the validity of results of clinical research and increases the feasibility of conducting meta-analyses. If the standards and procedures as outlined above are used, the use of PROs will contribute to improved patient care and clinical research.

© 2019 Elsevier Ltd. All rights reserved.

Introduction

Outcome in orthopedic pathology was traditionally determined based upon clinical data such as radiographic healing, rate of complications or secondary interventions, or range of motion (ROM). Professionals considered these data as objective. The patient perspective of treatment success (or failure) was not included in those outcomes. In the past decades, the use of the so-called Patient-Reported Outcomes (PROs) has increased drastically. This mimics the evolution from a paternalistic doctor towards shared decision making in daily practice; the patients opinion becomes more important. PROs are also referred to as subjective outcomes, 'soft' outcomes, or self-report measures.

PROs are generally defined as 'any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else' [1]. They include, among others, perceptions and opinions on symptoms, functioning, health-related quality of life (HRQoL), and satisfaction. PROs are increasingly integrated in clinical research,

but also a rapid growth in the use of PROs in daily clinical care and quality control of patient care is noted [2].

Most objective outcome measures are subject to interrater disagreement and often do not provide definitive answers about whether an intervention is useful from a patient's perspective [3]. Also, objective outcome measures may correlate poorly with a patient's perspective of his/her own feelings of wellness. Well-designed and rigorously tested PROs may be better validated and have greater reproducibility than objective outcomes [4] and even more importantly have more practical relevance for the patient.

This review aimed to provide a strategy to identify, select, and use PROs in orthopedic trauma studies and patient care. In addition to an overview of different types of PROs and their intended use, the selection process, requirements regarding quality and language, and reporting guidelines are discussed.

Types of PROs and their use

In general, PROs can be generic or condition-specific. Generic PROs assess the impact of musculoskeletal conditions on the overall health and well-being of a patient, and were designed for use in a broad range of conditions or interventions [5]. They often measure health-related quality of life (HR-QoL). HR-QoL is how a person's health affects his or her ability to carry out normal social

* Corresponding author.

E-mail address: e.vanlieshout@erasmusmc.nl (E.M.M. Van Lieshout).

and physical activities. Examples of commonly used HR-QoL instruments are the EuroQoL-5 D (EQ-5 D) [6,7], Short Form-36 (SF-36) or its shorter variant the Short Form-12 (SF-12) [8,9], and the Short Musculoskeletal Functional Assessment (SMFA) [10]. Generic PROs generally have favorable measurement properties, and can be useful for drawing comparisons about patient-reported health states across groups with varying clinical conditions [11]. However, the use of generic PROs in the outcome assessment of musculoskeletal conditions and interventions is sometimes questioned, as they fail to capture condition-specific functional outcomes. An exception to this is the SMFA, which has demonstrated good validity and contains questions related to musculoskeletal complaints [10]. The EQ-5 D has been advised as instrument of choice for health-economic evaluation in musculoskeletal research [12,13].

As opposed to generic PROs, region-specific (or anatomical site-specific) and condition-specific PROs claim to assess both symptoms and functional status in patient groups with specific musculoskeletal pathology. These PROs are targeted at specific injuries (e.g., Achilles tendon Total Rupture Score for Achilles tendon ruptures) [14], treatments (e.g. Western Ontario McMaster University Osteoarthritis Index (WOMAC) for patients with osteoarthritis who had arthroplasty of the knee or hip) [15], or anatomic regions (e.g., Disabilities of the Arm, Shoulder and Hand (DASH) and the Lower Extremity Functional Score (LEFS) for all sorts of injuries and conditions of the upper extremity and lower extremity, respectively [16,17]. Anatomic region-specific and condition-specific PROs include questions on aspects of health affected by a specific body region [18]. In theory they have great potential to evaluate domains of physical function and health-related quality of life commonly affected by the specified anatomical site [19]. Due to the focus on specific conditions or anatomic regions, the condition-specific PROs are often considered to be more sensitive to changes in health status than generic PROs [20].

Ideally, a PRO is completed by either the patient or a representative. However, several instruments used in musculoskeletal research are so-called mixed clinician-based functional outcome instruments. These combine questions answered by patients with results from physical examination or clinical tests performed by an outcome assessor. Examples of such mixed instruments are the Harris Hip Score [21] and the Constant–Murley score [22], which require range of motion measurements. Other mixed PROs may also include evaluation of the gait pattern, strength, or stability. Scoring systems that include physical examination or clinical tests are more prone to interobserver variability.

Poolman *et al.* recommended not to use mixed clinical and functional outcome instruments because of, among others, the risk of interobserver variability in the clinician-reported part of the instrument, [20]. They supported their recommendation by stating that validated outcome instruments that provide reproducible results are available. Poolman *et al.* instead encourage to use condition-specific instruments when available, as they are more sensitive to change.

Quality criteria for PROs

The question which PRO is the best is not easy to answer. It depends, among others, on the research question, the population of interest, the setting (e.g., research or patient care), and the infrastructure (e.g., paper or electronic versions). Similar as any outcome used in patient-centered outcomes research, PROs should meet the required quality standards.

The COSMIN checklist (COnsensus-based Standards for the selection of health status Measurement INstruments) was developed in an international Delphi study in order to evaluate the methodological quality of health status questionnaires [23]. These measurement properties include: content validity, internal consistency,

criterion validity, construct validity, reproducibility (consisting of agreement and reliability), responsiveness (also referred to as longitudinal validity), floor and ceiling effects, and interpretability [24]. Table 1 describes these quality criteria in more detail and lists the criterion for positive rating for each of the measurement properties [24].

A measurement tool is considered to be valid when it actually measures what it proposes to measure [24]. A poor quality PRO used in effectiveness studies may, for example, lead to unreliable and misleading results of the study, potentially resulting in harm to the patient or inappropriate use of resources. Any existing and newly developed PRO must have adequate measurement properties to detect small but relevant treatment effects or changes in individual and groups of patients [23–25].

Other important properties

Another important requirement is that the PRO is comprehensive for all patients (or healthy volunteers, as applicable). El-Daly *et al.* investigated the readability of 59 commonly used orthopedic-specific PROs that were available in English [26]. The Fleish Reading Ease Score (FRES) was determined; the authors consider a score of 79 points and over (out of 100 points), indicative for good readability by the average UK adult. They found that the mean score for all PROs investigated was 55 points (range 0–93), corresponding to text best understood by 16- to 18-year-old students. Seven of the 59 PROs had a FRES score between 79 and 93 points, corresponding to text that can be understood by the average UK adult. These included the generic PRO SF-36 [8] and the condition-specific PROs Oswestry Low Back Pain Score [27], the Lysholm Knee Score [28], and Modified Mayo Wrist Score [29]. The authors found that a surprising number of PROs were likely to be unreadable and potentially incomprehensible to most patients asked to complete them.

Finding and selecting the most appropriate PROs

Identifying and selecting the best PROs for your clinical research or monitoring of patients in clinical practice is a delicate process. The first step is to develop a detailed research question. Based on that, the next steps are (2) to determine what domains or constructs you want to measure (e.g., pain, physical function, or quality of life); (3) what population you aim to include (e.g., patients who have undergone a specific intervention or have a specific diagnosis); (4) to determine if an outcome measure exists, that captures all these domains for these patients; (5) to evaluate the psychometric evidence for the outcome measure(s) identified; and finally if the instrument is usable. Usability concerns items like the test burden, the time it takes to complete the questionnaire, whether or not the questionnaire can be used free of charge, and whether or not the scores are meaningful.

Identifying and selecting PROs typically starts with two systematic (literature) reviews. For a complete overview, the searches are best conducted in online general databases such as MEDLINE, Cochrane Library, EMBASE, and topic-specific databases such as the COSMIN database of systematic reviews of measurement instruments (<http://database.cosmin.nl/>).

In the first review, the available literature is searched for PROs that are commonly used in research focusing on injury or treatment of interest. From the resulting list of instruments, experienced musculoskeletal clinical researchers can identify which instruments would be the most suitable for evaluating the intervention under investigation and the targeted patient population.

Just because an instrument has been used frequently in previous studies does not guarantee its quality. The PROs used in your

Table 1

Definitions and quality criteria for measurement properties of health status instrument, according to the COSMIN checklist [23].

Property	Definition	Quality criterion for positive rating *
Content validity	The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire	A clear description is provided of the measurement aim, the target population, the concepts that are being measured, and the item selection AND target population and (investigators OR experts) were involved in item selection
Internal consistency	The extent to which items in a (sub)scale are intercorrelated, thus measuring the same construct	Factor analyses performed on adequate sample size (7 * # items and ≥ 100) AND Cronbach's alpha(s) calculated per dimension AND Cronbach's alpha(s) between 0.70 and 0.95
Criterion validity	The extent to which scores on a particular questionnaire relate to a gold standard §	Convincing arguments that gold standard is "gold" AND correlation with gold standard ≥ 0.70
Construct validity	The extent to which scores on a particular questionnaire relate to other measures in a manner that is consistent with theoretically derived hypotheses concerning the concepts that are being measured	Specific hypotheses were formulated AND at least 75% of the results are in accordance with these hypotheses
Reproducibility	The degree to which repeated measurements (test-retest) in steady populations provide similar answers	
Agreement	The extent to which the scores on repeated measures are close to each other (absolute measurement error)	MIC < SDC OR MIC outside the LOA OR convincing arguments that agreement is acceptable
Reliability	The extent to which patients can be distinguished from each other, despite measurement errors (relative measurement error)	ICC or weighted Kappa ≥ 0.70
Responsiveness (longitudinal validity)	The ability of a questionnaire to detect clinically important changes over time	SDC or SDC < MIC OR MIC outside the LOA OR RR > 1.96 OR AUC ≥ 0.70
Floor and ceiling effects	The number of respondents who achieved the lowest or highest possible score	$\leq 15\%$ of the respondents achieved the highest or lowest possible scores
Interpretability	The degree to which one can assign qualitative meaning to quantitative scores	Mean and SD scores presented of at least four relevant subgroups of patients and MIC defined

MIC, Minimal Important Change; SDC, Smallest Detectable Change; LOA, Limits of Agreement; ICC, Intraclass correlation; SD, standard deviation.

* The criteria for positive rating are shown. Further details on indeterminate and negative ratings are described in the original manuscript [23].

§ A gold standard is frequently unavailable, in those cases construct validity should be assessed instead.

research should meet the COSMIN standards for the targeted intervention and study population. The use of poor quality instruments may result in biased or unreliable effect estimates, and can thus potentially harm patients, mislead decision-makers relying on this evidence, and waste resources.

After identifying the PROs of interest, the second systematic literature review is aimed to make an overview of the measurement properties of these PROs. Using the COSMIN checklist, each measurement property is rated as positive, indeterminate (unclear from what is reported), or negative. If no information for a property is available in the literature, a rating of zero is given to indicate no evidence. If necessary, new primary research may be needed to focus on areas where the measure is either flawed or where no evidence is available. If none of the available PROs be valid, reliable, and responsive, development of a new PRO may be considered. Developing a new PRO is time-consuming and may easily take more than a year to complete or more depending on the complexity. After identifying a detailed aim of the instrument and the target population, it involves a stepwise approach that includes the following steps; (1) item generation (based on input from focus groups, literature, and a Delphi consensus process); (2) constructing a pilot questionnaire; (3) testing the pilot questionnaire in a sample of the target group (for reliability, validity, and ease of understanding); (4) generating a revised pilot questionnaire with subsequent testing in another sample of the target group; (5) eliminating items that perform poorly (mainly based upon reliability statistics); and (6) retesting the final questionnaire to determine all measurement properties as shown in Table 1 [30].

Many systematic reviews on the methodological quality of PROs are already available. Weel *et al.* evaluated Dutch foot-and ankle PROs [31] and reported that the Foot Function Index (FFI) and the Foot and Ankle Outcome Score (FAOS) [31] are considered the best instruments for foot and ankle complaints. Dacombe *et al.* conducted a systematic review on PROs for traumatic injuries to the hand and wrist used in randomized controlled trials [32]. The most commonly used PROs were the Disabilities of the Arm, Shoulder and Hand (DASH) questionnaire [16], Patient-Rated Wrist Evaluation (PRWE) [33], Gartland and Werley score [34], Michigan Hand

Outcomes score [35], Mayo Wrist Score [29], and Short Form-36 [8]. Of these, only the DASH and PRWE had evidence of reliability, validity, and responsiveness in patients with traumatic injuries to the hand and wrist. The other PROs either had incomplete evidence or evidence was gathered in a nontraumatic population [32]. Systematic reviews of the measurement properties are also available for other PROs used in orthopedic trauma, including the knee [36], total knee arthroplasty [37], anterior cruciate ligament injuries [38], total hip arthroplasty [39], upper extremity following trauma [40], elbow injuries [41], hand injuries [42,43] and many more musculoskeletal disorders and conditions [2].

Translation and cross-cultural adaptation of PROs

PROs are typically developed in one language, and subsequently translated in to other languages as needed. It is not sufficient to simply translate a previously validated instrument [44]. The American Academy of Orthopaedic Surgeons has adopted the Guideline for Cross Cultural Adaptation of Self-Report Measures, which states that translation and cultural adaptation is best done using a 5-step approach [45,46]. In step 1, two native speakers of the targeted language will individually translate the questionnaire. In step 2, *synthesis*, the translators and a panel of experts make combine the two translations into one. In step 3, the new version is translated back into English independently by two persons, native to the original language but highly experienced in the targeted language. In step 4, all translators of step 1 and 3 as well as a panel of experts combine the two back translated versions into one, and judge if this matches the original questionnaire. If necessary, adjustments to the new language version are made. This results in the (pre-)final version that is tested in step 5. The translated instrument needs to be revalidated before it can be applied in clinical or research settings.

Patients in different countries may respond differently to certain questions, irrespective of whether or not they speak the same language. Words can have subtle different meanings in different geographic locations. Therefore, it is very important to consider the population on which the outcome instrument was originally tested

[47]. If not properly validated in the population of interest, a validation study may be needed.

Reporting guidelines for PROs in study protocols and peer-reviewed manuscripts

Clinical studies that include PROs usually do not report any information on the instruments' measurement properties. Likewise, there is heterogeneity in data reporting across studies, both in the way the data are reported and in completeness of reporting (e.g., reported means may be inappropriate, and ranges are reported instead of standard deviations). Several reporting guidelines have been developed (<http://www.equator-network.org/library/>). A 15-item extension of the Standard Protocol Items – Recommendations for Interventional Trials (SPIRIT-PRO) was recently developed in order to guide the reporting of PROs in clinical trial protocols [48]. Similarly, a 5-item extension of the CONSolidated Standards Of Reporting Trials (CONSORT) for PROs was developed in order to guide the reporting of PROs used as primary or secondary outcome measures in randomized clinical trials (CONSORT-PRO) [49]. A task force of the International Society of Quality of Life (ISO-QOL) Research developed reporting standards for HRQOL outcomes in RCTs. [50]. In addition, ISOQOL developed a set of minimum standards for the design and selection of PROs for use in patient-centered outcomes research (PCOR) and comparative effectiveness research (CER) [51]. These may supplement the COSMIN criteria when choosing a PRO for use in research.

Core outcome measurement sets

In addition to standardizing the use and reporting of PROs, standardization of data collection for musculoskeletal research in general is needed. In 2014, Goldhahn *et al.* developed a Core Outcome Set (COS) for use in studies on distal radius fractures [52]. A COS is an agreed upon minimum selection of outcomes that should be measured and reported in all clinical trials for a particular health condition [53]. A COS increases the use of outcomes that are important to patients, clinicians, and researchers. It facilitates comparison across studies and ensures that the outcomes used have adequate measurement properties for their use. Finally, it should prevent selective outcome reporting like the reporting of only statistically significant findings. This improves the validity of results of clinical research and increases the feasibility of conducting meta-analyses [53–55].

The Core Outcome Measures in Effectiveness Trials (COMET) initiative was set up to assist and support researchers in their attempt to develop a COS [53]. In addition to the COS for distal radius fractures mentioned above [52], COSs are also available for other musculoskeletal diagnoses and interventions, including non-specific low back pain [56], knee pain after knee replacement [57], and hip fractures [58]. The COMET database currently contains over 1100 references of planned, ongoing and completed work (<http://www.comet-initiative.org/>).

Implementation of PROs in daily clinical routine

Individual physicians and hospitals increasingly use PROs, yet only few countries, such as England, Sweden, and the USA, have implemented the use of PROs on a regional or national level, [59]. In 2004, the National Institutes of Health (NIH) in the USA developed the Patient-Reported Outcomes Measurement Information System (PROMIS) together with an online assessment center to provide generic measures of the most important health outcome domains for many diseases [60,61]. PROMIS item banks assess physical (physical function, fatigue, sleep disturbance, sleep related

impairment, pain behavior, and pain interference), emotional (depression, anxiety, and anger), cognitive (applied cognition-abilities and applied cognition general concerns), and social health (ability to participate, satisfaction with social roles and activities, emotional support, instrumental support, informational support, and social isolation) [60–63]. PROMIS is freely available (www.assessmentcenter.net). Through the use of Computer Adaptive Testing (CAT), PROMIS health assessment data can be collected with minimal respondent burden.

Several initiatives to investigate clinicians' attitudes to using PROs in daily routine have been published (as referenced in [64]). Insight into the familiarity of trauma and orthopedic surgeons with PROs and their use in daily clinical routine is limited. Joeris *et al.* conducted a survey among 17,931 members of the AO Foundation (www.aofoundation.org; 9567 trauma surgeons, 6054 spine surgeons, and 2310 craniomaxillofacial surgeons) [64]. Of the responding trauma and orthopedic surgeons, 67% reported to be familiar with disease-specific PROs, 46% with generic PROs, yet only 27% use PROs in clinical practice. The results of this survey showed there is consensus on the usefulness of PROs, but implementation is limited. A majority of surgeons reported to be willing to implement PROs, if an adequate tool would be available.

Falavigna *et al.* conducted a survey among AO Spine members, aimed to assess barriers for full implementation of PROs [65]. The main barriers to implement PROs were lack of time to administer the questionnaires (57%) followed by lack of staff to assist in data collection (55%), and the long time to complete the questionnaires (46%). The authors concluded that 32% of spine surgeons did not use PROs routinely, which was attributed to a lack of knowledge regarding their importance, absence of reimbursement for this extra work, minimal financial support for clinical research, the cost of implementation, and lack of concern among physicians.

Future perspectives of PROs in musculoskeletal research and clinical practice

It is to be expected that the use of PROs in clinical practice and research will continue to increase. Increased use of PROs for musculoskeletal research requires more research on the measurement properties of PROMIS instruments in patients with orthopedic pathology. This requires rigorous systematic reviews of the existing evidence and new research on determining measurement properties of existing instruments. Developing a new PROs is only sensible if systematic review shows an obvious gap that cannot be solved using existing PROs. Whether or not PROMIS and Computer Adapted Testing are the ideal solution towards successful implementation of PROs in clinical routine and musculoskeletal research remains to be seen. Increased use of PROMIS requires the development of more item banks.

In addition to the above, developing new and implement existing Core Outcome Sets for many more musculoskeletal disorders will be the best approach to guide researchers. Having the same data for all studies on the same topic will be a huge step forward.

Conclusion

Generic and condition-specific PROs each have their specific pros and cons. Selection of instruments for a specific research question starts with a systematic literature review for identifying the mostly used instruments. From this list, the instruments that are most suitable for the intended use are selected. Readability, usability, and costs are considered at this stage. For the selected instruments, the literature is then searched systematically in order to determine if the measurement properties of these instruments meet the quality standard as outlined by the COSMIN group. Instruments need to be valid for the targeted condition,

intervention as well as population. If not available in the primary language of the targeted population, the instrument requires translation according to a state of the art procedure and validation in the new language. Reporting guidelines like the SPIRIT-PRO and CONSORT-PRO are available in order to guide the reporting of PROs in study protocol and outcome reports, respectively. Using Core Outcome Sets improves the validity of results of clinical research and increases the feasibility of conducting meta-analyses. If the standards and procedures as outlined above are used, the use of PROs will contribute to improved patient care and clinical research.

Ethics

Ethics Committee approval is not applicable to this review.

Funding statement

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. There are no statements to declare relating contributorship, data sharing, or ethics approval.

Declaration of Competing Interest

The authors (Esther M.M. Van Lieshout and Mathieu M.E. Wiffels) declare that they have no conflict of interests.

References

- https://www.qualityforum.org/Projects/n-r/Patient-Reported_Outcomes/Patient-Reported_Outcomes.aspx. Accessed January 08, 2019.
- Gagnier JJ. Patient reported outcomes in orthopaedics. *J Orthop Res* 2017;35(10):2098–108.
- Jackowski D, Guyatt G. A guide to health measurement. *Clin Orthop Relat Res* 2003;413:80–9.
- Pynsent PB. Choosing an outcome measure. *J Bone Joint Surg Br* 2001;83(6):792–4.
- Changulani M, Okonkwo U, Keswani T, Kalairajah Y. Outcome evaluation measures for wrist and hand: which one to choose? *Int Orthop* 2008;32(1):1–6.
- Brooks R, Rabin RE, Eds DC. The measurement and valuation of health status using EQ-5 D: a European perspective. Europe: Kluwer Academic Publishers; 2003.
- Lamers LM, Stalmeier PF, McDonnell J, Krabbe PF, van Busschbach JJ. [Measuring the quality of life in economic evaluations: the Dutch EQ-5 D tariff]. *Kwaliteit van leven meten in economische evaluaties: het Nederlands EQ-5 D-tarief*. *Ned Tijdschr Geneesk* 2005;149(28):1574–8.
- Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. conceptual framework and item selection. *Med Care* 1992;30(6):473–83.
- Ware J Jr, Kosinski M, Keller SD. A 12-Item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;34(3):220–33.
- Swiontkowski MF, Engelberg R, Martin DP, Agel J. Short musculoskeletal function assessment questionnaire: validity, reliability, and responsiveness. *J Bone Joint Surg Am* 1999;81(9):1245–60.
- McPhail SM, Bagraith KS, Schippers M, Wells PJ, Hatton A. Use of condition-specific patient-reported outcome measures in clinical trials among patients with wrist osteoarthritis: a systematic review. *Adv Orthop* 2012;2012:273421.
- Neugebauer E, Bouillon B, Bullinger M, Wood-Dauphinee S. Quality of life after multiple trauma—summary and recommendations of the consensus conference. *Restor Neurol Neurosci* 2002;20(3–4):161–7.
- Van Beeck EF, Larsen CF, Lyons RA, Meerding WJ, Mulder S, Essink-Bot ML. Guidelines for the conduction of follow-up studies measuring injury-related disability. *J Trauma* 2007;62(2):534–50.
- Nilsson-Helander K, Thomee R, Silbernagel KG, Thomee P, Faxen E, Eriksson BI, et al. The Achilles tendon total rupture score (ATRS): development and validation. *Am J Sports Med* 2007;35(3):421–6.
- Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988;15(12):1833–40.
- Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The upper extremity collaborative group (UECG). *Am J Ind Med* 1996;29(6):602–8.
- Binkley JM, Stratford PW, Lott SA, Riddle DL. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. North American orthopaedic rehabilitation research network. *Phys Ther* 1999;79(4):371–83.
- Wells GA, Russell AS, Haraoui B, Bissonnette R, Ware CF. Validity of quality of life measurement tools—from generic to disease-specific. *J Rheumatol Suppl* 2011;88:2–6.
- Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. *Med Care* 1989;27(3 Suppl):S217–32.
- Poolman RW, Swiontkowski MF, Fairbank JC, Schemitsch EH, Sprague S, de Vet HC. Outcome instruments: rationale for their use. *J Bone Joint Surg Am* 2009;91:41–9 Suppl 3.
- Harris WH. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. An end-result study using a new method of result evaluation. *J Bone Joint Surg Am* 1969;51(4):737–55.
- Constant CR, Murlley AH. A clinical method of functional assessment of the shoulder. *Clin Orthop Relat Res* 1987;214:160–4.
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The Cosmin study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63(7):737–45.
- Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60(1):34–42.
- Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use, 5th ed. Oxford: Oxford University Press; 2008. p. 399.
- El-Daly I, Ibraheim H, Rajakulendran K, Culpan P, Bates P. Are patient-reported outcome measures in orthopaedics easily read by patients? *Clin Orthop Relat Res* 2016;474(1):246–55.
- Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980;66(8):271–3.
- Lysholm J, Gillquist J. Evaluation of knee ligament surgery results with special emphasis on use of a scoring scale. *Am J Sports Med* 1982;10(3):150–4.
- Amadio PC, Berquist TH, Smith DK, Ilstrup DM, Cooney WP. 3rd, Linscheid RL. Scaphoid malunion. *J Hand Surg Am* 1989;14(4):679–87.
- Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. *BJA Education* 2017;17(4):137–44.
- Weel H, Zwiers R, Siersevelt IN, Haverkamp D, van Dijk CN, Kerkhoffs GM. [Dutch-language patient-reported outcome measures for foot and ankle injuries: a systematic review] *Nederlandstalige patientgerapporteerde uitkomstmaten voor voet- en enkelaandoeningen; een systematisch overzicht*. *Ned Tijdschr Geneesk* 2015;159:A8831.
- Dacombe PJ, Amirfeyz R, Davis T. Patient-reported outcome measures for hand and wrist trauma: is there sufficient evidence of reliability, validity, and responsiveness? *Hand (N Y)* 2016;11(1):11–21.
- MacDermid JC, Turgeon T, Richards RS, Beadle M, Roth JH. Patient rating of wrist pain and disability: a reliable and valid measurement tool. *J Orthop Trauma* 1998;12(8):577–86.
- Gartland JJ Jr, Werley CW. Evaluation of healed colles' fractures. *J Bone Joint Surg Am* 1951;33-A(4):895–907.
- Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan hand outcomes questionnaire. *J Hand Surg Am* 1998;23(4):575–87.
- Wang D, Jones MH, Khair MM, Miniaci A. Patient-reported outcome measures for the knee. *J Knee Surg* 2010;23(3):137–51.
- Gagnier JJ, Mullins M, Huang H, Marinac-Dabic D, Ghambaryan A, Eloff B, et al. A systematic review of measurement properties of patient-reported outcome measures used in patients undergoing total knee arthroplasty. *J Arthroplasty* 2017;32(5):1688–97 e7.
- Gagnier JJ, Shen Y, Huang H. Psychometric properties of patient-reported outcome measures for use in patients with anterior cruciate ligament injuries: a systematic review. *JBJS Rev* 2018;6(4):e5.
- Gagnier JJ, Huang H, Mullins M, Marinac-Dabic D, Ghambaryan A, Eloff B, et al. Measurement properties of patient-reported outcome measures used in patients undergoing total hip arthroplasty: a systematic review. *JBJS Rev* 2018;6(1):e2.
- Dowrick AS, Gabbe BJ, Williamson OD, Cameron PA. Outcome instruments for the assessment of the upper extremity following trauma: a review. *Injury* 2005;36(4):468–76.
- The B. Reininga IH, El Moumni M, Eygendaal D. Elbow-specific clinical rating systems: extent of established validity, reliability, and responsiveness. *J Shoulder Elbow Surg* 2013;22(10):1380–94.
- Schoneveld K, Wittink H, Takken T. Clinimetric evaluation of measurement tools used in hand therapy to assess activity and participation. *J Hand Ther* 2009;22(3):221–35 quiz 36.
- van de Ven-Stevens LA, Munneke M, Terwee CB, Spauwen PH, van der Linde H. Clinimetric properties of instruments to assess activities in patients with hand injury: a systematic review of the literature. *Arch Phys Med Rehabil* 2009;90(1):151–69.
- McKenna SP, Doward LC. The translation and cultural adaptation of patient-reported outcome measures. *Value Health* 2005;8(2):89–91.
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)* 2000;25(24):3186–91.
- Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993;46(12):1417–32.
- Johnson JA, Luo N, Shaw JW, Kind P, Coons SJ. Valuations of EQ-5 D health states: are the United States and United Kingdom different? *Med Care* 2005;43(3):221–8.

- [48] Calvert M, Kyte D, Mercieca-Bebber R, Slade A, Chan AW, King MT, et al. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: the spirit-pro extension. *JAMA* 2018;319(5):483–94.
- [49] Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD, et al. Reporting of patient-reported outcomes in randomized trials: the consort pro extension. *JAMA* 2013;309(8):814–22.
- [50] Brundage M, Blazeby J, Revicki D, Bass B, de Vet H, Duffy H, et al. Patient-reported outcomes in randomized clinical trials: development of Isoqol reporting standards. *Qual Life Res* 2013;22(6):1161–75.
- [51] Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISO-QOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res* 2013;22(8):1889–905.
- [52] Goldhahn J, Beaton D, Ladd A, Macdermid J, Hoang-Kim A. Distal radius working group of the international society for fracture R, et al. Recommendation for measuring clinical outcome in distal radius fractures: a core set of domains for standardized reporting in clinical practice and research. *Arch Orthop Trauma Surg* 2014;134(2):197–205.
- [53] Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. *Trials* 2012;13:132.
- [54] Kirkham JJ, Gargon E, Clarke M, Williamson PR. Can a core outcome set improve the quality of systematic reviews?—a survey of the co-ordinating editors of cochrane review groups. *Trials* 2013;14:21.
- [55] Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;340:c365.
- [56] Chiarotto A, Deyo RA, Terwee CB, Boers M, Buchbinder R, Corbin TP, et al. Core outcome domains for clinical trials in non-specific low back pain. *Eur Spine J* 2015;24(6):1127–42.
- [57] Wylde V, MacKichan F, Bruce J, Goberman-Hill R. Assessment of chronic post-surgical pain after knee replacement: development of a core outcome set. *Eur J Pain* 2015;19(5):611–20.
- [58] Haywood KL, Griffin XL, Achten J, Costa ML. Developing a core outcome set for hip fracture trials. *Bone Joint J* 2014;96-B(8):1016–23.
- [59] Black N. Patient reported outcome measures could help transform healthcare. *BMJ* 2013;346:f167.
- [60] Broderick JE, DeWitt EM, Rothrock N, Crane PK, Forrest CB. Advances in patient-reported outcomes: the NIH PROMIS(R) measures. *EGEMS (Wash DC)* 2013;1(1):1015.
- [61] Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The patient-reported outcomes measurement information system (PROMIS): progress of an NIH roadmap cooperative group during its first two years. *Med Care* 2007;45(5 Suppl 1):S3–S11.
- [62] DeWalt DA, Rothrock N, Yount S, Stone AA, Group PC. Evaluation of item candidates: the PROMIS qualitative item review. *Med Care* 2007;45(5 Suppl 1):S12–21.
- [63] Gershon RC, Rothrock N, Hanrahan R, Bass M, Cella D. The use of PROMIS and assessment center to deliver patient-reported outcome measures in clinical research. *J Appl Meas* 2010;11(3):304–14.
- [64] Joeris A, Knoll C, Kalampoki V, Blumenthal A, Gaskell G. Patient-reported outcome measurements in clinical routine of trauma, spine and craniomaxillo-facial surgeons: between expectations and reality: a survey among 1212 surgeons. *BMJ Open* 2018;8(6):e020629.
- [65] Falavigna A, Dozza DC, Teles AR, Wong CC, Barbagallo G, Brodke D, et al. Current status of worldwide use of Patient-Reported Outcome Measures (PROMs) in spine care. *World Neurosurg* 2017;108:328–35.